FUZZY INFERENCE SYSTEM FOR AUTOMATED IMAGE-BASED KARYOTYPING

Aliaksandr Kurachkin[1], Vasiliy Sadau[1]

[1]Faculty of Radiophysics and Computer Technologies, Belarusian State University, Belarus
lawliet29@gmail.com

Karyotyping, or chromosome analysis, is a series of steps to produce karyotype information (i.e. the number of chromosomes and their structure), usually based on visual metaphase analysis. A normal human karyotype contains 22 diploid autosomes and 2 gender chromosomes for a total of 46 chromosomes, each of them with specific structure and size, the most common reference being International System for Human Cytogenic Nomenclature (ISCN). Chromosome aberrations - structure disordering, chromosome fragment displacement and abnormal number of chromosomes - usually indicate an error in cell division cycle, and can be detected by cytogeneticist by ground-truth reference comparison during karyotyping. Early detection of such abnormalties is extremely important in prenatal diagnostics, as well as cancer cell analysis and treatment planning [1].

Chromosome analysis is commonly performed by manual visual expert analysis, although there are several well-known algorithms that allow automated chromosome separation and classification from slide shots. The input images for such algorithms are captured using high-magnification electron microscope and contain an image of a glass slide with individual chromosome species extracted from cells arrested during metaphase.

The expected output of an automated chromosome analysis system is a karyogram - a set of separated, ordered and classified chromosomes extracted from the image, that can be examined individually. The reference comparison for each chromosome is usually based on chromosome length, the configuration of alternating brighter and darker stripes along the chromosome length and their brightness distribution, presence of sattelites, and centromere location.

The most challenging problems are chromosome separation and classification. Separation is a process of extracting parts of the slide image that are only relevant to a single chromosome, independently of the others present on the image. However, chomosome species present on the slide often overlap, requiring a precise boundary detection and cutout analysis. Even when boundaries of a single chromosome are detected correctly, some parts of the separated chromosome that belong to one of the other species (i.e. because of the overlap) can alter further analysis. Chromosome classification is the procedure of assigning a specific chromosome number to the separated chromosome.

Existing classification methods are usually performed via sequential analysis of an individual separated chromosome and expectation maximization techniques based on a specific encoding of brightness pattern alternations. However, these algorithms are usually very sensitive to individual non-structural chomosome deformations like bending, and generally perform worse when the observed species overlap percentage is significant.

The proposed solution to this problem is to enhance existing sequential analysis algorithms with the introduction of a fuzzy inference system for classification [2]. The basis of fuzzy inference analysis is higher-order Denver chromosome classification, which groups individual chromosomes into 7 classes based on their centromere position. Relative centromere detection can be performed by iterative chromosome species analysis and is far less sensitive to non-structural defromations. Chromosome brightness distribution and relative length can be used to further refine Denver class selection.

The input terms for fuzzy inference system are centromere position, brightness distribution as mean and standard deviation and relative length, calculated directly from separated chomosome. Centromere position fuzzy set values include sets for metacentric, submetacentric and acrocentric chromosomes with sigmoidal activation functions. Brightness input variable uses a regular triangular 7-level grid for mean and standard deviation. Finally, relative chromosome length is provided as a crisp, non-fuzzy value. Fuzzy inference rules bind input parameter values to their respective output Denver classification classes. For inference itself, Mamdani algorithm is used.

Resulting Denver classification projection can be used to enhance initial classification estimate of traditional algorithms and increase their accuracy, especially for cases of non-structural chromosome species deformations and overlap.

[1] A. P. Britto, G. Ravindran, A Review of Cytogenetics and its Automation, Journal of Medical Sciences, vol 7 (1), pp. 1-18 (2007).
[2] T. J. Ross, Fuzzy logic with engineering applications, 3rd ed. John Wiley & Sons, Ltd., 607p., (2010).